CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION CFSTI
DOCUMENT MANAGEMENT BRANCH 410.11

LIMITATIONS IN REPRODUCTION QUALITY

ACCESSION # AD 603808

- [ ] 1. WE REGRET THAT LEGIBILITY OF THIS DOCUMENT IS IN PART UNSATISFACTORY. REPRODUCTION HAS BEEN MADE FROM BEST AVAILABLE COPY.
- [x] 2. A PORTION OF THE ORIGINAL DOCUMENT CONTAINS FINE DETAIL WHICH MAY MAKE READING OF PHOTOCOPY DIFFICULT.
- [ ] 3. THE ORIGINAL DOCUMENT CONTAINS COLOR, BUT DISTRIBUTION COPIES ARE AVAILABLE IN BLACK-AND-WHITE REPRODUCTION ONLY.
- [ ] 4. THE INITIAL DISTRIBUTION COPIES CONTAIN COLOR WHICH WILL BE SHOWN IN BLACK-AND-WHITE WHEN IT IS NECESSARY TO REPRINT.
- [ ] 5. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED, DOCUMENT WILL BE AVAILABLE IN MICROFICHE ONLY.
- [ ] 6. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED DOCUMENT WILL NOT BE AVAILABLE.
- [ ] 7. DOCUMENT IS AVAILABLE IN MICROFICHE ONLY.
- [ ] 8. DOCUMENT AVAILABLE ON LOAN FROM CFSTI (TT DOCUMENTS ONLY).
- [ ] 9.

PROCESSOR: eal

TSL-107-10/64

603808

X

COPY 1 of 1 COPIES

1

SOME TESTS OF THE RANDOMNESS OF A MILLION DIGITS

Bernice Brown

P-44

Ocotober 19, 1948

Approved for OTS release

22 p $1.00 hc
$0.50 mf

603808

DDC
RECEIVED
AUG 19 1964
DDC-IRA C

The RAND Corporation
· SANTA MONICA · CALIFORNIA

(20)

SUMMARY

The randomness of a table of a million digits produced at Project Rand by a random digit generator was examined by applying four tests. The four tests were: (1) Frequency test, (2) Poker test, (3) Serial test and (4) Run test. The complete table of a million digits was subjected to the first two tests. The last two tests were applied to a sample of 50,000 digits from the table. All computations were accomplished by means of I.B.M. equipment.

There was no evidence of any unusual divergence from the theoretical expectations in any of the tests. It would be difficult to construct a series of digits which would succeed in evading all four of these tests. The examination of the million digits did not reveal any bias. There was nothing to indicate that the digits were not being produced with equal probabilities.

Introduction. This report is concerned with an examination of the randomness of a million digits which were produced at Project Rand by a random digit generator. This machine is basically a random noise generator, a pulse counter and a recorder. The generator was designed to produce pulses at an average rate of the order of 10^5 pulses per second. These pulses operate an electronic binary counter. At intervals of one second, this counter was 'read' electrically. The count was converted by means of relays into the decimal system and the last digit recorded automatically on I.B.M. cards. A report describing the machine will be forthcoming. A series of tests applied to periodic samples of digits produced by the machine indicated that a small bias (.0006) favoring the odd digits might be present in the table of a million digits. Following a suggestion made by G. W. Brown, this table was re-randomized by a process that involved forming sums of paired digits modulo 10. It is the purpose of this report to present the results of an examination of the million re-randomized digits in an attempt to insure the randomness of the numbers for the user of random numbers in common types of sampling inquiries.

Procedure. Four methods of examining the million digits for randomness have been utilized. All tables and computations were accomplished by use of I.B.M. equipment.

1. Frequency Tests. A table of aggregate frequencies was prepared with a record of the total frequency of occurrence of each of the ten digits for one million numbers in 20 samples of 50,000 digits each.

The million digits were also examined in blocks of 1000 digits each. A set of 1000 summary cards was prepared; each card containing the ten observed frequencies; the computed Chi-square for these frequencies and the probability of getting χ^2 equal to or smaller than that observed. These quantities were calculated from the formulae,

$$\chi_i^2 = \frac{1}{100} \sum_{j=0}^{9} (x_{ij} - 100)^2 \qquad (i=1,2,\ldots,1000)$$

$$P(\chi^2) = \frac{1}{\Gamma(\frac{9}{2})} \int_0^{\chi_i^2} \left(\frac{t^2}{2}\right)^{\frac{7}{2}} e^{-\frac{t^2}{2}} d\left(\frac{t^2}{2}\right)$$

where x_{ij} represents the observed frequency of the digit j in the ith sample of 1000 digits. The distribution of the one thousand calculated Chi-squares was compared with the theoretical χ^2 distribution.

2. Poker Test. The million digits were scanned in groups of five digits each, simulating 200,000 poker hands and a record made of the number of hands which contained five differing digits, busts (symbol abcde), number of pairs (symbol aabcd), two pairs (symbol aabbc), three digits alike (symbol aaabc),

<u>full house</u> (symbol aaabb), <u>four digits alike</u> (symbol aaaab) and <u>all five digits alike</u> (symbol aaaaa). Records of the frequencies in the seven classes of poker hands were kept for every 1000 poker hands examined. The tabulation sheets thus supplied us with 200 frequency distributions, each of which could be compared with the theoretical frequency distribution.

3. <u>Serial Test</u>. The serial test was applied to a sample of 50,000 digits from the table of the one million re-randomized digits. A ten by ten table was prepared showing the frequency with which any one digit was followed by any other digit. This block of digits constitutes a 5 percent sample from the million digit table. The object of the serial test was to see if a given digit tended to be associated with any other digit.

4. <u>Run Test</u>. The run test was applied to the same sample of 50,000 digits which was subjected to the serial test. The occurrence of two adjacent like digits in the series is defined as a run of length 2. For example, if

$x_{ij} = x_{i+1,j} \left({i=1, \ldots 50{,}000 \atop j=0, \ldots 9} \right)$ and $x_{i-1,j}$ and $x_{i+2,j} \neq x_{ij}$ then x_{ij} and $x_{i+1,j}$ constitute a run of length 2 in digit j. Similarly, if $x_{ij} = x_{i+1,j} = x_{i+2,j}$ and $x_{x-1,j} \neq x_{ij}$, $x_{i+3,j} \neq x_{ij}$ then a run of length 3 has occurred in digit j. A run of length

3 was not counted as two runs of length 2 but was independent. The last digit in the series $x_{50,000,j}$ was considered as being adjacent to the first number $x_{1,j}$ in order to complete the circuit. The sample of 50,000 digits was assumed to be a random sample from an infinite universe in which one digit was as likely to occur as another.

Results

1. Frequency Tests. The frequency of the occurrence of the ten digits in each of 20 successive samples of 50,000 is shown in Table I, together with total frequencies. For $N = 10^6$, the expected mean (m) is 10^5 with standard deviation (σ) of 300. Six of the observed frequencies are within the interval $m \pm \sigma$ and in only two cases is the deviation from expected more than 2σ. The χ^2 value for 9 degrees of freedom is 13.3 and the probability of exceeding this value is approximately 0.15.

The total number of even digits is 500,586 as against 499,414 odd digits. Under the hypothesis that an even digit is as likely to occur as an odd, the probability of a departure as great as 586 from an even division is 0.24. A difference greater than this might occur about one time in four. Hence the deviation does not appear to be significant.

The frequency Chi-square for 9 degrees of freedom may be analyzed into three component parts; one degree of freedom for the comparison of odd and even digits, four degrees of freedom

for variation between frequencies within the group of odd digits and four degrees of freedom between digits within the even number group. These components of Chi-square are as follows:

	χ^2	d.f.	Probability
Comparison of frequencies of odd vs. even digits.	1.37	1	0.25*
Frequencies within group of odd digits.	7.90	4	0.10
Frequencies within group of even digits.	4.04	4	0.40

The result of this analysis of Chi-square into its components upholds the hypothesis that the sampling is from a population in which the digits occur with equal probabilities. There is no indication of any differentiation in the behavior of odd and even digits. An inspection of the frequencies shows that the digit which appeared most frequently (two) was associated with a probability of 0.1006 while the least frequent digit (nine) had a probability of 0.0993.

The computed sample χ^2 values are shown in the column on the right for samples of 50,000 digits. Each of these

---

* Probability of a greater χ^2 arising by chance.

TABLE I

Frequencies of One Million Digits (Re-randomized)

No.	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	χ^2
1	4923	5013	4916	4951	5109	4993	5055	5080	4986	4974	7.556
2	4870	4956	5080	5097	5066	5034	4902	4974	5012	5009	10.132
3	5065	5014	5034	5057	4902	5061	4942	4946	4960	5019	6.078
4	5009	5053	4966	4891	5031	4895	5037	5062	5170	4886	15.004
5	5033	4982	5180	5074	4892	4992	5011	5005	4959	4872	13.846
6	4976	4993	4932	5039	4965	5034	4943	4932	5116	5070	7.076
7	5011	5152	4990	5047	4974	5107	4869	4925	5023	4902	14.116
8	5003	5092	5163	4936	5020	5069	4914	4943	4914	4946	13.051
9	4860	4899	5138	4959	5089	5047	5030	5039	5002	4937	13.410
10	4998	4957	4964	5124	4909	4995	5053	4946	4995	5059	7.212
11	4948	5048	5041	5077	5051	5004	5024	4886	4917	5004	7.142
12	4958	4993	5064	4987	5041	4984	4991	4987	5113	4882	6.992
13	4968	4961	5029	5038	5022	5023	5010	4988	4936	5025	2.162
14	5110	4923	5025	4975	5095	5051	5035	4962	4942	4882	10.172
15	5094	4962	4945	4891	5014	5002	5038	5023	5179	4852	16.261
16	4957	5035	5051	5021	5036	4927	5022	4988	4910	5053	4.856
17	5088	4989	5042	4948	4999	5028	5037	4893	5004	4972	5.347
18	4970	5034	4996	5008	5049	5016	4954	4989	4970	5014	1.625
19	4998	4981	4984	5107	4874	4980	5057	5020	4978	5021	6.584
20	4963	5013	5101	5084	4956	4972	5018	4971	5021	4901	6.584
Total	99802	100050	100641	100311	100094	100214	99942	99559	100107	99280	13.316
$f_{ij}-10^5$	-198	50	641	311	94	214	-58	-441	107	-720	

Chi-square values is associated with 9 degrees of freedom. The distribution of the observed Chi-squares is in line with expectation. Two of the 20 values give probabilities of exceeding the observed χ^2 value of more than 0.90; two give probabilities of less than 0.10. In 12 samples the χ^2 values are small enough to give probabilities of more than 0.50; 8 samples show probabilities of less than 0.50. For $n = 50,000$ the expected mean is 5,000 with standard deviation of $30\sqrt{5}$, or 67.08. Of the 200 recorded frequencies, 129 or 64.5 percent are within one standard deviation of the mean which compares favorably with the 68 percent of the theoretical distribution. There are eight cases or 4 percent in which the deviation of observed frequency from expected frequency exceeds 2σ.

The calculation of a sample χ^2 value for 9 degrees of freedom from each sample of 1000 digits offered an opportunity for an examination of the Chi-square distribution. In Figure 1, the cumulative distribution function of χ^2 has been plotted against the χ^2 values. The 1000 χ^2 values were arranged in 50 class intervals with probability in each class approximating

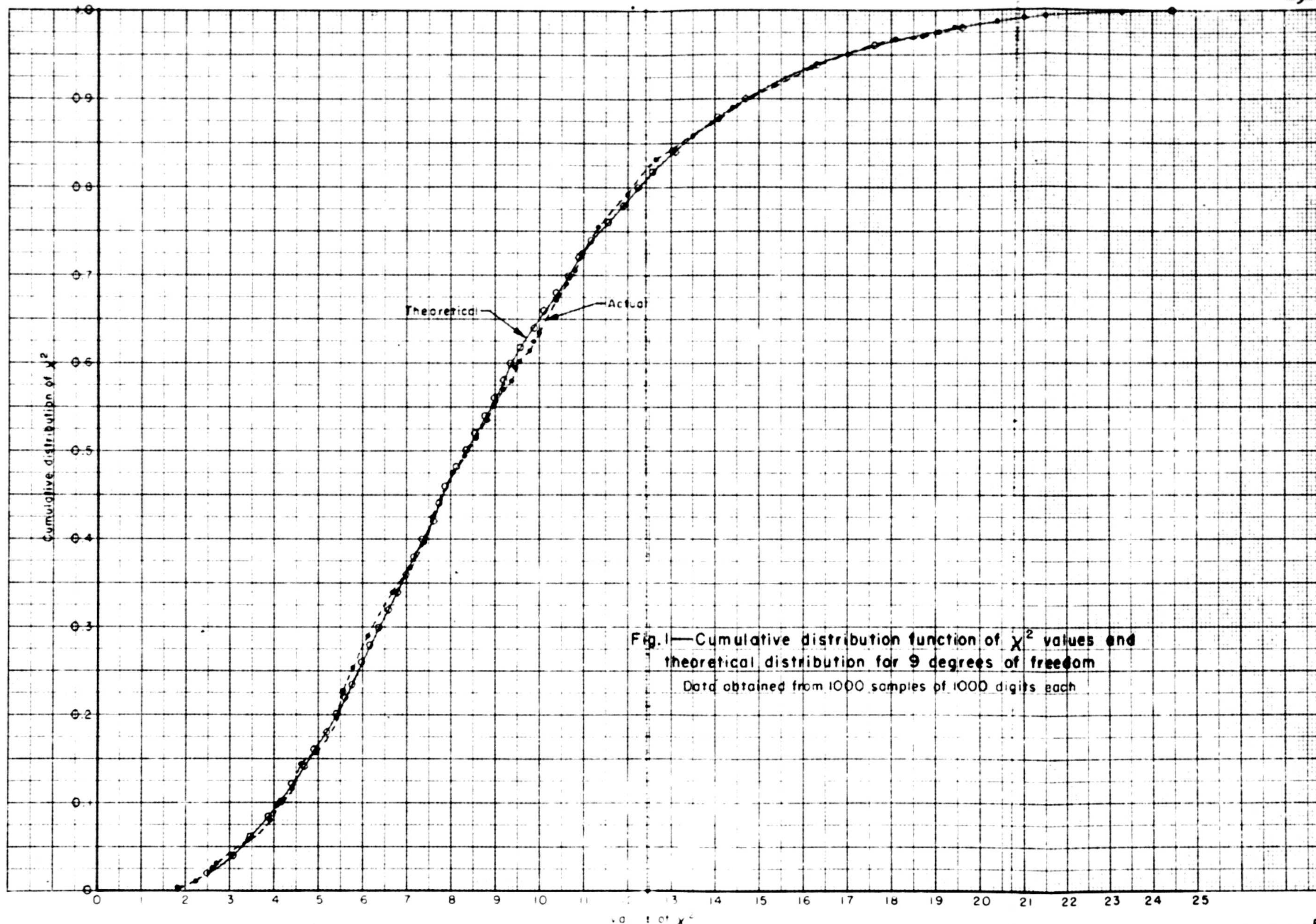

Fig. 1—Cumulative distribution function of χ^2 values and theoretical distribution for 9 degrees of freedom

Data obtained from 1000 samples of 1000 digits each

2 percent.* It will be observed that the theoretical and actual distributions are in agreement except for a slight disparity in the range of values from 9.0 to 10.5. A χ^2 test was applied to determine whether the observed frequencies in the 50 classes differed significantly from the theoretical values. A χ^2 of 54.6 was obtained. There seems to be nothing unusual about getting a χ^2 as high as 54.6. Since it is associated with 49 degrees of freedom, the resulting probability is about 0.55 using Fisher's χ^2 approximation.

2. <u>Poker Test</u>. Two hundred Chi-squares were calculated which measured the dispersion between expected and observed frequencies in each sample of 1000 poker hands. The expected distribution of frequencies for samples of 1000 are <u>busts</u> 302.4, <u>pairs</u> 504, <u>two pairs</u> 108, <u>three</u> 72, <u>full house</u> 9, <u>fours</u> 4.5 and <u>fives</u> 0.1. The expected frequencies in the last two classes were small, hence were combined into one group.

The observed distribution of χ^2 is compared with the theoretical distribution for five d.f. in Table II.

---

* Wald and Mann [1] have shown that an optimal way of determining the number of intervals k in which the range of the sample χ^2's is to be divided, is given by $k = 4\sqrt[5]{\frac{2(N-1)^2}{c^2}}$ where c is the value of the normal deviate at the level of significance desired and N is total frequency. $k \sim 50$ was determined from this relationship.

TABLE II

Frequency Distributions of Chi-square Values

Probability	Values of χ^2	Expected Frequency	Observed Frequency
$P > .90$	0 - 1.60	20	22
$.90 > P > .80$	1.61 - 2.35	20	19
$.80 > P > .70$	2.36 - 3.00	20	22
$.70 > P > .60$	3.01 - 3.70	20	19
$.60 > P > .50$	3.71 - 4.35	20	20
$.50 > P > .40$	4.36 - 5.20	20	29
$.40 > P > .30$	5.21 - 6.10	20	22
$.30 > P > .20$	6.11 - 7.30	20	15
$.20 > P > .10$	7.31 - 9.20	20	15
$P < .10$	9.21 or more	20	17
		200	200

$$\chi^2 = 7.7 \text{ for } 9 \text{ d.f.} \quad P = 0.55$$

The agreement of the two distributions indicates that the Chi-squares obtained experimentally from the table of random digits are compatible with the tabular χ^2 values for 5 degrees of freedom.

The frequencies obtained from the poker test were combined and the results of the aggregate of 200,000 poker hands are shown in Table III.

TABLE III

Poker Test on Million Digits (2000,000 Poker Hands)

Classes	Expected Frequency	Observed Frequency
Busts (abcde)	60,480	60,479
Pairs (aabcd)	100,800	100,570
Two pairs (aabbc)	21,600	21,572
Threes (aaabc)	14,400	14,659
Full house (aaabb)	1,800	1,788
Fours (aaaab)	900	914
Fives (aaaaa)	20	18
	200,000	200,000

$\chi^2 = 5.5$ for 5 d.f. P = 0.35

The agreement of observed and expected frequencies in Table III is good. The only sizable difference is in the class of threes (aaabc). The difference, 259 is 2.25 times the theoretical value of the standard deviation, 115.6 for this class. Considered by itself, a deviation of this magnitude would ordinarily occur one time in 40. The value of χ^2 for all the frequencies of 5.5 with 5 degrees of freedom is not excessive.

The frequencies of Table III may be broken down into ten equal groups, each representing 20,000 poker hands (100,000

digits) in order to look at the consistency of the behavior of the poker hands. The theoretical mean and standard deviation for each class was calculated under the assumption of equal probability of occurrence of each digit. The comparison of actual and theoretical values of mean and standard deviation is shown in Table IV.

TABLE IV

Mean and Standard Deviation of Frequencies in 7 Classes of Poker Hands

Classes	Theoretical Mean	Actual Mean	Theoretical Std. Dev.	Actual Std. Dev.
abcde	6048	6047.9	64.9	60.3
aabcd	10080	10057.0	70.7	78.4
aabbc	2160	2157.2	43.9	45.8
aaabc	1440	1465.9	36.9	26.6
aaabb	180	178.8	13.4	8.9
aaaab	90	91.4	9.5	11.5
aaaaa	2	1.8	1.4	1.9

With the possible exception of the class of poker hand in which three of a kind appeared (aaabc), the actual means and standard deviations are satisfactory estimates of the theoretical values. In the class referred to above, the mean was overestimated and the standard deviation was underestimated.

3. <u>Serial Test</u>. The frequency of the occurrence of all possible pairs of digits is given in Table V.

TABLE V

Serial Test

Frequency of Occurrence of 1st Digit

Frequency of Occurrence of 2nd Digit	0	1	2	3	4	5	6	7	8	9	Total
0	508	510	451	500	513	475	494	508	463	501	4923
1	456	514	523	472	561	490	486	512	503	496	5013
2	509	474	493	476	481	527	491	454	475	536	4916
3	507	514	484	466	485	507	483	498	514	493	4951
4	502	504	502	513	526	493	525	550	520	474	5109
5	489	481	466	478	513	481	504	533	544	504	4993
6	471	496	514	540	485	489	530	516	514	500	5055
7	504	486	506	513	510	512	539	504	491	515	5080
8	488	507	493	530	524	465	513	485	520	461	4986
9	489	527	484	463	511	554	490	520	442	494	4974
Total	4923	5013	4916	4951	5109	4993	5055	5080	4986	4974	50000

This table was formed by entering a pair of digits $x_i x_j$ in the ith column and the jth row; e.g., if a sequence of four digits is 6203, a frequency is recorded in the 6th column, 2nd row - one in 2nd column, 0 row and one in 0 column, 3rd row. In order to complete the circuit and make the 50,000 digits give exactly 50,000 pairs, the last digit was combined with the first

to form a pair.

Let us consider, first of all, the probability of getting the sample from a population in which the digits were equally likely to occur. The frequency χ^2 as calculated from the row or column totals is 7.56. This is based on 9 degrees of freedom. The probability of getting a Chi-square equal to or greater than 7.56 is about 3 in 5. This sample of 50,000 does not show unusual divergence. The frequencies obtained here would be encountered often in sampling.

Three Chi-square tests were applied to the data in Table V. The first tested the hypothesis that the array was a random sample from a population in which one pair of digits was as likely to occur as another. The expected frequency in each cell was 500. (Kendall and Smith [3], apply this test to Tippett's published table of random numbers). The χ^2 value for 90 degrees of freedom was 107.8. The probability of getting a $\chi^2 > 107.8$ for 90 degrees of freedom is about 0.10. There is no evidence that the sample was not drawn from the population specified.

The second tested the hypothesis that any given digit was as likely to be followed by one digit as by another. The method of computing the expected number in each cell is based on the actual frequency: e.g. since the actual frequency of

zeros was 4923, the expected value in each cell in the 1st column of Table V was 492.3. A similar computation was made for the other nine frequencies. The χ^2 value was 98.9. The probability of getting a $\chi^2 > 98.9$ with 90 d.f. is 0.25. The deviations from expected values are not large enough to discredit the hypothesis.

The third test was used to check the hypothesis that any given digit is as likely to be preceded by one digit as by another. If the preceding digit is equally likely to be 0,1,2,...9, the expected value in each cell of the first row was 492.3, in the 2nd row 501.3 etc. The value of χ^2 was 100.4. The probability of getting a $\chi^2 > 100.4$ is approximately 0.20. The evidence presented by the sample is compatible with the stated hypothesis.

4. Run Test. The expected frequency of runs in the sample of 50,000 digits was computed from formulas in a paper on the distribution theory of runs given by Mood [4]. The theoretical frequencies of runs of length r are given in Table VI, together with the actual frequencies observed in the sample.

TABLE VI

Run Test

Length of Run	Expected Frequency	Observed Frequency
r = 1	40500	40410
r = 2	4050	4055
r = 3	405	421
r = 4	40.5	48
r = 5	4.5	5

It is clear that the observed distribution is in good agreement with the theoretical distribution. There is no evidence that the number and length of runs observed in this sample were not compatible with a sample from an infinite universe in which one digit was as likely to occur as another.

Conclusion. The examination of the million re-randomized digits has shown no discrepancies other than the expected sampling fluctuations. The frequency test indicated that all the digits occurred approximately an equal number of times. The distribution of sample Chi-square values agreed closely with the theoretical distribution. When the digits were examined in groups of five each, the observed frequencies compared favorably with the expected frequencies. The serial test revealed no tendency for any particular pair of digits to be

associated together. The number of runs observed in the sample and the distribution of their lengths conformed to the run theory advanced. None of the tests contradicts the assumption of randomness in the series of one million re-randomized digits.

REFERENCES

[1] Wald, A. and Mann, H. B., Annals of Mathematical Statistics, Vol. XIII, No. 3, September 1942.

[2] Fisher, R. A., ''Statistical Methods for Research Workers'', 10th Edition, p. 81 and 92.

[3] Kendall, M. G. and Smith, B. B., Journal Royal Statistical Society 101 (1938), Supplement (1939).

[4] Mood, A. M., Annals of Mathematical Statistics, Vol. XI, 4, December 1940.